



## De l'usage des bases de données INA par le chercheur

Jérôme Bourdon, Marie-Françoise Lévy, Cécile Méadel

### ► To cite this version:

Jérôme Bourdon, Marie-Françoise Lévy, Cécile Méadel. De l'usage des bases de données INA par le chercheur. Les dossiers de l'audiovisuel, 1994, 54, pp.80-82. halshs-00192685

**HAL Id: halshs-00192685**

**<https://shs.hal.science/halshs-00192685>**

Submitted on 30 Nov 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **De l'usage des bases de données INA par le chercheur**

Jérôme Bourdon<sup>1</sup>, Marie-Françoise Levy<sup>2</sup>, Cécile Méadel<sup>3</sup>

### **1. De la spécificité des bases de données de l'INA**

Les bases de données ouvrent à deux types de documentation, primaires et secondaires. Ce second type est le plus souvent utilisé par les chercheurs pour leurs recherches bibliographiques, pour sortir des articles ou des ouvrages, pour faire un état des travaux ou confronter leurs recherches à celles d'autrui. Les bases de données de l'INA relèvent elles du second type, elles leur donnent accès, non à des travaux, mais à des informations brutes, à des fiches sur des émissions de radio et de télévision. Elles sont l'équivalent d'un répertoire ou d'un catalogue d'archives spécialisées, mais, dans l'ensemble, beaucoup plus riche d'informations qu'un simple catalogue.

Considérons d'abord la base Imago (qui repère les émissions de télévision postérieures à 1974). Pour chaque émission, Imago nous indique le genre (au sens usuel du mot, qui fait l'objet aussi d'une définition précise dans la base), le générique, le nom des participants, la durée, la chaîne et le jour de diffusion (mais non l'heure), le matériel technique disponible, l'origine (juridique) de la production. Les émissions d'actualité (journaux et magazines) sont décrites plan par plan, en vue d'une réutilisation immédiate par les professionnels (qui détruit parfois le fil d'un reportage). Pour les autres émissions, dites "de production", on dispose d'un résumé. Ce résumé a été réalisé "en direct", par un documentaliste-spectateur, au moment de la diffusion. Il prend donc la forme d'un "récit" qui permet de restituer, par exemple dans un débat, les temps forts, le point de vue des participants.

Mior (la base de l'avant 1974) est plus complexe: il s'agit d'une reprise informatique d'un fichier papier, qui indique généralement titre, date et heure de diffusion, chaîne, et participants. L'analyse du contenu est beaucoup plus lacunaire, réduite à des indications de séquences ou de participants. En outre, Mior est disponible pour les émissions de production, mais non pour l'actualité, sauf cas de sous-catalogage thématique réalisés à l'occasion de telle ou telle production (on dispose par exemple d'une saisie informatique des sujets magazines portant sur la Guerre d'Algérie).

Enfin, INASON (la base radio) est la plus hétérogène. Les fonds sont éclatés en plusieurs séries (Toulouse, etc.). Enfin, à l'inverse de ce qui se produit en télévision, les

---

<sup>1</sup> INA, direction de la recherche

<sup>2</sup> Institut d'Histoire du Temps présent, CNRS

<sup>3</sup> Centre de Sociologie de l'Innovation, Ecole des Mines de Paris

documents les plus récents, donc les plus facilement accessibles, sont les plus mal décrits par la base, qui contient, au contraire, des informations très riches sur le début du fonds (les années cinquante).

COMPLÉTER CECILE?

NOte G.PIEJUT: prévoir fac-similé de fiches ici.

## **2. Des usages "professionnels" aux usages "scientifiques"**

Les bases de données de l'INA ont été construites pour répondre aux exigences des professionnels de la radio-télévision. Les informations retenues sont tournées d'abord vers cette exploitation; c'est pour cela, par exemple que l'on trouve en bonne place des informations sur la nature des supports, sur les droits de diffusions... La description même des documents est axée vers la réexploitation et a suivi ses aléas, la notion même de réexploitation professionnelle ayant fortement évolué : dans les débuts, on essayait de préserver et de mémoriser quelques plans pittoresques. Aujourd'hui, la télévision se recyclant elle-même sans cesse, la description prolifère, notamment en actualité, puisque tout peut re-servir (tout invité peut devenir célèbre, tout fait peut être objet de réexploitation).

Ces finalités professionnelles posent-elles problème aux chercheurs? Il faut distinguer deux aspects dans cette question.

Le premier est important mais contingent et peut-être provisoire. Il porte que les modalités d'accès à ces bases. L'outil d'interrogation est construit en fonction des questions des professionnels et il est mal armé pour répondre aux demandes des chercheurs. Ainsi, Imago renacle lorsqu'on lui demande de traiter des corpus un peu volumineux. Autre point: le langage d'interrogation est obscur, hermétique au non-documentaliste, fort peu accueillant. Cela peut ralentir dans une certaine mesure la recherche, mais cela ne peut l'empêcher. En outre, l'INA travaille à établir un langage d'interface rendant plus "friendly" le contact avec ses bases de données.

Le second point pourrait être plus sérieux mais il nous semble déboucher sur un cul de sac: la description professionnelle des documents est-elle contradictoire avec des descriptions de recherche? En fait, lorsque l'on interroge les chercheurs sur la manière dont ils souhaiteraient voir indexer des documents audiovisuels, on se rend compte que les descriptifs prolifèrent mais surtout qu'ils divergent considérablement. Il n'y a pas aujourd'hui consensus sur le langage d'indexation du document audiovisuel. En fait, on ne sait pas pour l'heure ce qu'est une exploitation scientifique des documents audiovisuels - et, en tout état de cause - on ne saurait en arrêter définitivement la définition. Certes, des règles élémentaires de critiques des sources exigent que l'on soit renseigné sur l'origine du document, la date et le créneau de diffusion. Au-delà, il y a

tout ce que la base n'analyse pas - et qu'on ne peut pas énumérer car on ne peut prévoir comment vont s'orienter les intérêts, ou, plus profondément, comment vont évoluer les découpages cognitifs. Ce n'est jamais le rôle d'une base de données de se substituer à l'imagination scientifique ou professionnelle. Par exemple une recherche sur l'évolution de la voix (débit, intensité) et, indissociablement, de la prise de son, pour une analyse de la parole publique ne trouverait que des indications rarissimes dans les bases, il y a pourtant là une piste importante de recherche.

En outre, une telle question oublie -ou dénie- tout l'intérêt de la base comme outil professionnel fort utilisé. Les rubriques d'Inason ou de Mior nous donnent des indications précieuses sur les intérêts de hommes de radio et de télévision à un moment daté de leur histoire. Aucune base de donnée ne s'abstrait de son époque mais là, elles nous initient directement vers les préoccupations des milieux professionnels.

Ceci dit, tenons nous en aux préoccupations dominantes de la recherche. Certains sujets sont plus facilement abordables, d'autres non. Tentons donc un classement des "qualités" (au sens neutre du terme) de la base en fonction des démarches.

### **3. Les démarches de recherche, du quantitatif au qualitatif.**

Partons d'une démarche ancienne, encore fréquemment employée, inspirée des premières analyses de presse : les "analyses du contenu", au sens technique du terme: découpage du document en unités mesurables (et identifiées de la même façon par des chercheurs différents), traitement statistique. On cherche à répondre à des questions du type: quelle place la télévision consacre t-elle à un public (émissions pour enfants), à un thème, à une personnalité, à un événement, selon les époques?

Les bases de l'INA sont mal équipées pour ce type d'analyse: elles ne décrivent pas l'ensemble de la grille, elles excluent les émissions dont l'INA n'est pas, en tout ou en partie, propriétaire (mais ceci est destiné à évoluer avec la mise en place du dépôt légal). Elles ne permettent pas des traitements qui prennent en compte des calculs de durées. On peut simplement dénombrer les unités appartenant à un genre télévisuel, ou traitant d'un certain "thème". Ceci donne des ordres de grandeur, tributaire de l'évolution des langages et des pratiques d'indexation. Le sujet "cinéma" est fréquemment évoqué dans des émissions "omnibus" (magazines de l'après-midi), pour des durées très variables, dans des sujets qui ont été indexés ou non, selon les moments. En interrogeant le mot clef "cinéma" dans tous les champs de la base, on aura donc un ordre de grandeur du volume d'émissions considérés, permettant d'amorcer une recherche, en aucun cas un élément d'analyse. De surcroît, la base est peu appropriée au traitement de grandes quantités de documents. Pour pallier cet inconvénient, pour éviter aussi de se trouver face à des masses chiffrées considérables, la base permet de procéder à des interrogations par chaînes, par époque, par genre. En combinant ces critères, on peut

procéder à d'utiles sondages, plutôt de que tenter immédiatement une interrogation globale.

Sur ces exemples, on saisit immédiatement que l'interrogation sur des thèmes doit servir moins à quantifier leur traitement par la télévision qu'à repérer des émissions qui en ont traité. Loin de toute souci quantitatif, on sera donc emmené à multiplier les interrogations pour approcher toutes les émissions concernées. Si l'on s'intéresse aux immigrés, par exemple, on devra interroger les mots-clefs "minorités", "étrangers", plus les différentes nationalités concernés: l'interrogation donne alors des indications sur les émissions et sur le lexique descriptif dominant à une époque.

Outre les thèmes et les participants, on peut aussi approcher les émissions à partir de leur genre. Mais cette catégorie professionnelle, qui a beaucoup évolué, est d'un usage délicat. Prenons trois exemples: le genre "journal télévisé" a peut-être évolué, mais les émissions de ce type sont clairement identifiées - pour l'essentiel. Le cas du documentaire est plus ambigu (frontières avec le magazine ou le reportage). Enfin, pour des catégories nouvelles de programmation, (les "reality show" ou la "télévision vérité"), la base ne peut pas (et d'un point de vue documentaire, ne doit sans doute pas) "courir après" le lexique professionnel. Les émissions hybrides sont donc classé dans la catégorie magazine, avec des mentions de genres comme "justice", "police" - mais qu'en est-il, par exemple, d'une émission comme "Perdu de vue" (justice? recherche de disparu?).

Reste que la base est très précieuse pour faire l'historique d'un genre (en restreignant peut-être à une personne donnée). Elle met à jour l'oublié (considérable) de la programmation. A cet égard, les émissions oubliées, les essais, les séries avortées, sont légions. Elles obligent à de nouvelles interrogations.

Plutôt que le thème ou le genre, on pourra s'intéresser à des catégories moins ambiguës, moins sujettes aux fluctuations conceptuelles. Cas d'interrogation univoque: les personnalités présentes dans les émissions sont un excellent moyen d'interroger la base, qui permet de réaliser des "télébiographies" (historique et évolution des apparitions télévisées). On peut aussi systématiquement vérifier quelle personnalité sont associées entre elles, quels rôles leur sont données, de quoi elles parlent pour l'essentiel.

L'on peut enfin imaginer de croiser des interrogations portant sur ces différentes séries de données qui précèdent: d'essayer par exemple d'établir des corrélations entre des genres d'une part, des intervenants, des thèmes d'autre part.

#### **4. Des bases de données au document**

Sans qu'on l'ait clairement distingué jusque là, on s'est toujours placé dans la position du chercheur qui utilise la base pour repérer des documents auxquels il demande d'avoir accès. Or, pour répondre à certains catégories de questions (celles précisément qu'analyse la base), on peut utiliser directement la base. L'analyse des supports

techniques, l'apparition de certains participants et le rôle qu'on leur fait tenir, le genre et la forme d'une émission (ces derniers les réserves que l'on a déjà émises), sont, pour certaines de leurs caractéristiques, possibles à partir des seules bases. Pourtant les chercheurs auront presque toujours besoin de compléter leurs informations par le recours aux archives écrites, à la documentation et en particulier à la presse de programmes, aux entretiens avec les acteurs. On peut par exemple imaginer des réunions et des débats avec d'anciens téléspectateurs sur la mémoire des émissions.

La tentation peut être grande d'utiliser la base comme un "équivalent de l'antenne". Elle ne peut jouer ce rôle. En tout état de cause, même en travaillant sur des données que l'on tirerait seulement des bases de données, on aura intérêt à sortir un échantillon de documents de différentes époques : le ton, la mise en scène, le dispositif, peuvent affecter le traitement du thème, la présentation du témoin, rendre hétérogène ce que l'on croyait homogène, et, plus simplement, renvoyer le chercheur à la vérité phénoménale de l'émission, qui ne "dit" jamais rien sur un thème ou un intervenant, mais est, d'abord et toujours, spectacle et mise en scène. Bref, la base n'est pas transparente aux documents.